CHAPTER

5.14

Digital Television Audio Loudness Management

JIM STARZYNSKI
NBC Universal
New York, NY

BACKGROUND AND HISTORY

Along with high-definition (HD) pictures, digital television (DTV) introduced cinema-quality surround sound that had a tremendous increase in audio dynamic range compared to its analog predecessor. With this increase in dynamic range came the unintentional consequence of severe loudness variation when loudness was not managed correctly. Initially, the home listening DTV audience often felt the negative impact of the new audio system in the form of extreme loudness variation between commercials and programs on a given channel and loudness variation when changing channels across the dial. Viewers were not accustomed to the inconvenience of constantly adjusting the volume with their remote controls, but were now burdened with having to do so to enjoy digital HDTV.

At the onset of the analog-to-digital transition and at the crux of the problem, TV station engineers, whose priority and typical focus is on keeping signals on the air, were now, among many other pressing demands, presented with DTV's entirely new Dolby AC-3 audio system. The new Advanced Television System Committee (ATSC) DTV standard used Dolby AC-3 digital audio encoding to reduce the bit rate needed to deliver up to 5.1 channels of high-bandwidth, high dynamic range surround sound. This new AC-3 digital encoding system was a completely new and relatively complicated technology, and it shared little similarity with loudness-control techniques previously used with analog NTSC transmission. Station engineers still needed to become familiar with requirements for correct operation, including spending the time to master the system's enhanced capabilities and establishing matching program and advertising content delivery specifications.

In the meantime, suppliers were continuing to mix their digitally recorded soundtracks to specifications that were intended for use with analog transmission systems that would constrain range by design, with no regard for DTV's extreme range capabilities. No new rules were in place for correctly delivering DTV mixes, and no guidelines existed to help understand how to effectively use DTV's expanded range for creative advantage. In this environment, uncontrolled soundtracks compounded by incorrectly set station encoding often resulted in loudness swings of 10 dB or more.

Meanwhile, in the mid-2000s, affordable rear-projection and flat-panel televisions had entered the market. As a result the DTV audience was growing significantly, and so was viewer annoyance with poor DTV loudness control. As the situation worsened and viewer complaints grew, television engineers became as frustrated as the audience.

By 2006, the ATSC had recognized the problem, and in 2007 it took action by creating an audio group composed of industry experts to identify technical issues and to write guidelines for solving the problem. Shortly thereafter (and unaware of the ATSC's work in progress), the U.S. Congress responded to the American public's dissatisfaction and initiated HR 1084, *The Commercial Advertisement Loudness Mitigation Act* ("CALM Act"). The original bill required the audio of TV advertisements that accompanied programming on U.S. television stations to not be "noisy or strident" and "not be presented at modulation levels higher than the program material they accompany." In 2010, a revised version of the CALM Act became law, mandating the use of newly released ATSC's loudness practices by the U.S. television broadcasting industry.

This chapter on digital television loudness management provides background and explains proven, effective and *required* practices for mastering loudness across the television audio ecosystem, with particular focus on television engineering and operations.

The Analog-to-Digital Transition: A New DTV Audio System and the Makings of a Growing Problem

At the commencement of the analog-to-digital television transition, an additional 6 MHz television channel was assigned to each broadcaster to facilitate the transition from the legacy NTSC analog service to a new, forthcoming and mandated digital television service using the ATSC standard. The latter would eventually remain as the broadcaster's only transmitted channel once the date for the analog shutoff was reached. This situation posed the interim operation of an additional digital broadcast plant, with content that, for the most part, mirrored the output of the analog plant (i.e., it was a simulcast transmission). To accomplish this, standard-definition (SD) digital video signals were upconverted in many cases to high definition, and encoded for emission (using MPEG-2 video compression) using the ATSC standard on the station's DTV channel.

Analog audio signals were digitized and encoded to the ATSC's standard using the Dolby AC-3 system (branded as *Dolby Digital* in the consumer electronics world), requiring a unique set of essential parameters in the encoder to describe the audio in a completely new DTV audio metadata authoring process. One such parameter was *channel mode*. At the commencement of DTV, most if not all audio soundtracks were encoded in 2/0 (stereo), given that the source of the audio was content from the analog, dual-channel mono or stereo-only NTSC channel used by North American (and other) broadcasters. This is where DTV audio's similarity with analog ends.

The Introductiovn of Digital Audio Encoding

At the time, DTV audio encoding was performed by an onboard audio encoder that resided alongside the video section of the ATSC broadcast encoder, or by a stand-alone, stereo Dolby DP567 digital audio encoder. The output stream of the device was connected to the ATSC encoder in audio pass-through mode, enabling the 567's AC-3 audio stream to remain intact during emission encoding. The connection was straightforward, but the setup for both types of encoders could be puzzling due to new metadata authoring requirements, which bore little or no resemblance to the analog TV audio emission setup. The DP567 stereo encoder was quickly supplanted by the DP569, a 5.1 audio channel capable unit, which was for many years thereafter the only device available to encode digital, discrete surround sound for DTV.

As seen in Figure 5.14-1, the DP569 bears little physical resemblance to earlier analog audio equipment, and its setup was equally new to the user. As part of the metadata authoring process, numerous parameters required examination of their default settings by the broadcaster. Some settings were for informational purposes, but many directly controlled the audio being encoded.

As listed in Table 5.14-1, "dialog level" (which would come to be known as "dialnorm" or "dialog normalization") tops the list of metadata control parameters. This item would become the focal point of loudness control for digital television for content creation, broadcast emission and, most importantly, the listener experience, and it remains so today. That being just one of twenty-seven new parameters, it's easy to understand how the broadcast operator could become consumed with many new settings yet still need to investigate the overall workings of DTV loudness control using this very important dialnorm parameter. This all took place in the context of those engineers having to focus on the revenue-producing analog NTSC channel while establishing a new ATSC DTV channel, as well.

In NTSC analog television, audio loudness was controlled to ensure that 100% modulation of the frequency-modulated FM audio carrier was not exceeded. Analog TV broadcasting is capable of ~50 dB of range, with dialog typically 17 dB below 100% modulation.[1] ATSC AC-3 audio is capable of >100 dB of dynamic range; perceptually, this equates to 32 times the dynamic range of analog broadcast audio. See Figure 5.14-2.

NTSC audio was frequently processed with dynamic range control ahead of the studio-transmitter-link or at the transmitter. This was often performed by a stand-alone device, or with an optional card in or alongside the stereo generator at the transmitter. This device was relied upon for protection against overmodulation and, as a by-product, it was effective against loudness shifts while evolving to create pleasant although *altered* sound for the listener. As a result, this process changes the sound from what the content creator intended, but it was necessary to fit the audio within the modulation constraints of the analog transmission. See Figure 5.14-3.

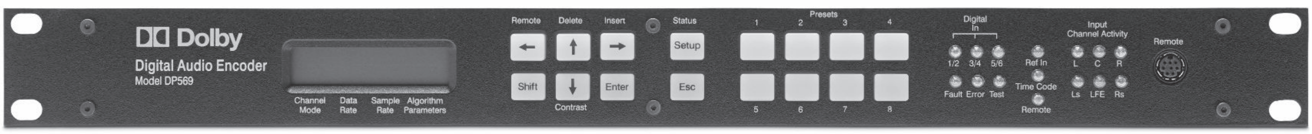

FIGURE 5.14-1 The Dolby 569 Digital Audio Encoder, the first device capable of encoding 5.1 surround sound for ATSC Digital Television.

TABLE 5.14-1
The Metadata Parameter table from the Dolby DP569 Digital Audio Encoder Operator's Manual. (Extended Bitstream Information parameters are in italics.)

Metadata Parameter	Informational	Control
Dialogue Level		×
Channel Mode		×
LFE Channel		×
Bitstream Mode	×	
Line Mode Compression		×
RF Mode Compression		×
RF Overmodulation Protection		×
Center Downmix Level		×
Surround Downmix Level		×
Dolby Surround Mode		×
Audio Production Information	×	
Mix Level	×	
Room Type	×	
Copyright Bit	×	
Original Bitstream	×	
Preferred Stereo Downmix		×
Lt/Rt Center Downmix Level		×
Lt/Rt Surround Downmix Level		×
Lo/Ro Center Downmix Level		×
Lo/Ro Surround Downmix Level		×
Dolby Surround EX Mode		×
A/D Converter Type	×	
DC Filter		×
Lowpass Filter		×
LFE Lowpass Filter		×
Surround 3 dB Attenuation		×
Surround Phase Shift		×

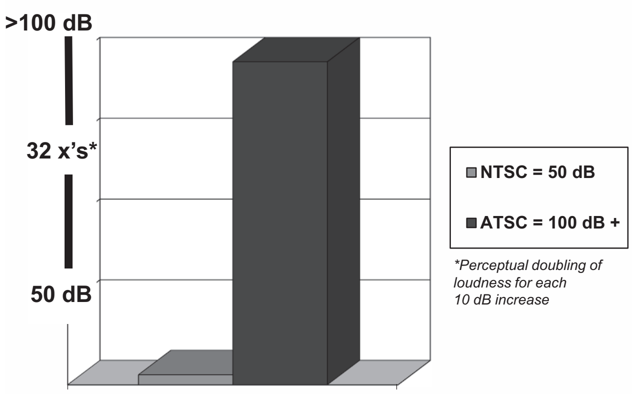

FIGURE 5.14-2 Analog versus digital TV audio dynamic range.[1]

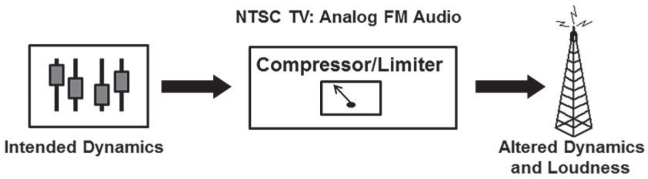

FIGURE 5.14-3 Typical NTSC audio loudness control.

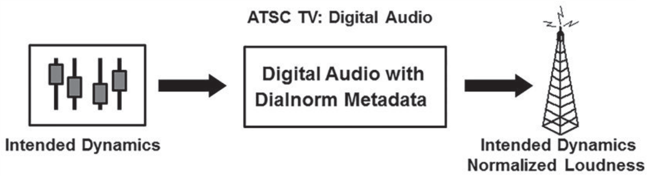

FIGURE 5.14-4 ATSC audio loudness control, in which intended audio dynamics are preserved in the broadcast via dialog normalization.

As mentioned earlier, the AC-3 audio system defined in the ATSC digital television standard uses metadata to control loudness and other audio parameters more effectively without permanently altering the dynamic range of the content.[2] The *dialnorm* parameter is used to normalize the loudness of the content. As required in the standard, for correct operation of the loudness normalization system, the value of the dialnorm parameter must match the measured loudness of the content. This has come to be called "the Golden Rule."[3] See Figure 5.14-4.

Analog TV audio measurement techniques did not focus on the actual *loudness* of the content. Familiar volume unit (VU) and peak program meters (PPM) measured electrical (not perceived loudness) levels; they were used to protect equipment against overload and to establish a nominal operating point above the noise floor while maintaining adequate headroom below clipping. In contrast, two key elements for audio measurement in the digital TV system are *true peak* and *loudness* levels. True peak measure helps the operator protect against clipping, and a useable loudness measurement is needed to avoid unwanted wide variations in what the listener hears.[4] The difference between the familiar, available measurement techniques used for analog TV and the new ones required for DTV added yet another challenge for the operator.

Introduction of the Broadcast Loudness Meter

This DTV loudness paradigm included the development of a new digital television broadcast loudness meter, whose first of many incarnations would be the Dolby LM100 (see Figure 5.14-5). At its introduction, the device read in units of *Leq(A)*, an available legacy technique effective for measurement of motion picture soundtracks. The meter lacked true peak measurement capability at the time. Leq(A) would soon be supplanted by ITU-R BS.1770: "Algorithms to Measure Audio Program Loudness and True Peak Audio Level." AuQ28

FIGURE 5.14-5 The Dolby LM100 Broadcast Loudness Meter.

This seminal document is described in Chapter 5.18 and will be further discussed later in this chapter.

The meter introduced a selectable algorithm called Dialogue Intelligence, allowing the LM100 to automatically base ITU-R BS.1770–1 or Leq(A) measurements on the portions of the input signal that contain the characteristics of dialog.[5] The LM100 did not use a familiar sweeping needle or LED cluster display. It measured and displayed its readout as a single numeric value of the audio loudness, measured over a period of time. The operator could choose either "short-term," the report of an updated average of the content loudness for the previous 10-second period, effective for measurement for live mixing in real time, or "infinite," an ongoing average that could be controlled by pause/reset selections on the front panel, and effective for measuring recorded content. In addition, the LM100 was capable of reading the dialnorm parameter from an AC-3 bitstream, and it displayed that value next to the measured audio loudness reading for easy comparison. The LM100 was limited to reporting loudness of 2 channels of PCM sources and up to 5.1 channels of Dolby AC-3. It could also measure up to 8 channels of Dolby E, which was at the time a new, mezzanine-level[6] audio codec[7] intended for professional postproduction and distribution use.

Equipped with a broadcast loudness meter, the operator could now measure the standardized loudness of content at production, postproduction and ingest, as well as the loudness of the digital television station output and its dialnorm value at transmission.

Content Loudness and Dialnorm

Equipped with loudness metering but with no new content delivery recommendations for audio loudness in place, operators quickly realized that programs and commercials were being recorded and delivered at widely variable loudness levels. To fulfill DTV's promise of theater-like sound, no compressor-limiter-type control devices were in the path to manage useable range and normalize the loudness at the station output. Instead, it was intended that the operator rely on dialnorm to normalize the loudness. To do so as designed, dialnorm metadata, reflecting the measured loudness of the audio, needed to accompany each piece of content, travelling from content creation, through distribution and to emission encoding. If present, a single dialnorm value could dynamically change within the AC-3 stream to match the measured average loudness of complete shows or complete commercials it accompanied. Decoders would apply the specific dialnorm value at content boundaries, attenuating the audio at the output of the receiver by an offset value needed to achieve the system normalization reference value of –31.

For example, content encoded with a –20 dialnorm value would be attenuated 11 dB, while content encoded with a –31 value would not be attenuated. If dialnorm always matched actual content loudness (the Golden Rule), the audience would be presented with smooth loudness transitions at content boundaries and channel changes, all normalized to –31 in a process that was seamless, while maintaining the content creator's artistic intent.

Encoding Dialnorm Metadata with Content

The means to include dialnorm metadata with the audio essence at this time was limited to encoding the audio at postproduction with Dolby E—or in a subsequently developed process, as vertical ancillary (VANC) data—inside the serial digital interface (SDI) video and audio signal.[8]

The Dolby E codec was developed to expand the number of audio channels available at production from the limits of four-channel VTRs and servers to eight channels, therefore facilitating DTV's new 5.1 surround format and more.[9] The codec included metadata authoring compatible with AC-3. This data and audio was efficiently designed to be distributed over a single AES3 pair,[10] either alone or embedded in SDI. Unfortunately, this would add another layer to DTV audio challenges by adding two frames of delay and requiring special monitoring gear. Though very effective when operated in a managed system, and with some success in postproduction situations, terrestrial broadcasters shied away from Dolby E's universal usage. VANC metadata insertion practices came later on, but existed under two different standards. They were rejected by the Hollywood postproduction audio community due to the risk posed to a show master when assembling the data into the video portion of a tape, an unfamiliar technique for audio operators.

The Results of Wide-Range DTV Audio without Loudness Control

In many cases, with no valid loudness metadata and no analog loudness control type device in place, audio was delivered to the audience unaltered and without loudness normalization, exactly as supplied by various content creators, with the ability to access the full >100 dB of DTV's dynamic range. In addition, the broadcast encoder's dialnorm value was typically left at its default value of –27, and this value seldom reflected the actual loudness of the content.

As evidence of operators' growing dissatisfaction with the performance of the system, a number of stations changed their static dialnorm value to –31, essentially "turning off" the DTV loudness normalization system. Exacerbating the problem, this now created excessive loudness variations on channel changes by users, in addition to shifts at content boundaries within a channel (e.g., programs vs. commercials). In addition, stations attempting to apply the Golden Rule were now penalized in some markets, because a station adhering to the ATSC standard and the Golden Rule had its audio attenuated to AC-3's –31 loudness reference, while stations turning off the system and broadcasting audio at a typical –20 dBFS were transmitting audio 11 dB louder than compliant stations.

As affordable flat-panel displays hit the market in the mid-2000s and the DTV audience grew, so did annoyance with television audio. The ineffectiveness of the system irked the listener with its loudness variations often exceeding 10 dB, and it plagued the broadcaster in the form of growing user complaints and impending government regulation.

The need for a reevaluation of the system or a mutually agreed-upon industry recommended practice was never more apparent.

ATSC Takes Action

With growing dissatisfaction over the use and performance of the DTV audio loudness system, television station engineers referred to the ATSC standard for guidance on exactly what was required to manage audio loudness.

What they found in the original ATSC A/53 Audio Standard was the following:

> *Dialogue Level*
>
> The value of the dialnorm parameter in the AC-3 elementary bit stream shall indicate the level of average spoken dialogue within the encoded audio program. Dialogue level may be measured by means of an "A" weighted integrated measurement (LAeq). (Receivers use the value of dialnorm to adjust the reproduced audio level so as to normalize the dialogue level.)[11]

TV engineers questioned the wording of the standard. Why was average spoken dialog used as the reference? What if content had no dialog? Was Leq(A) (corrected abbreviation for "LAeq" that appeared in the standard) adequate for the measurement of TV sound?

Fortunately, work was underway at this time in the ITU-R (International Telecommunications Union—Radiocommunication Sector) on development of a contemporary loudness measurement recommendation. Aside from that, however, there was no other industry-wide recognized and suitable information, and no other standards or official recommended or best practices that could be used at the time. The section on dialog level of the ATSC standard seemed ambiguous and unfamiliar. In the meantime, consumers did not expect and were annoyed by large changes in audio loudness from program to interstitials, and from channel to channel. Clearly, a proper solution was required.[12]

Formation of ATSC S6–3 Audio Subgroup

In December 2006, the ATSC Board of Directors recognized the loudness problem and assigned a work item to the responsible Technology Group. In April of 2007, ATSC S6–3 formed an ad hoc audio subgroup to develop a scope of work to "investigate issues relating to variations in reproduced audio levels within the ATSC Digital Television System," with an objective to have their work "ultimately result in the drafting of a Recommended Practice on Audio Loudness."[13]

The proper expertise and industry representation was needed in order for the group's work to be effective and for its efforts to achieve consensus and approval by the ATSC membership and by the industry at large. With this in mind, a group of engaged and assertive audio experts from commercial TV and cable networks, public television, professional audio equipment manufacturers, academia and other standards organizations was assembled. The group's expertise included broadcast system and laboratory engineering, audio production and postproduction mixing, system and circuit design, motion picture sound, and cable, satellite and telco television distribution, along with input from other standard development and industry trade organizations. With these key stakeholders in place, the group took on the task of solving the DTV audio loudness problem.

The first face-to-face meeting of the group was a "Digital Television Audio Loudness Summit" held in July 2007 at the University of Southern California (USC), attended by over 60 participants. Its intent was to create a heightened awareness of the issues involved. The outcome of the meeting concluded that DTV loudness was indeed a serious problem shared by many parts of the industry along with the TV audience.

ATSC S6–3 then determined that a thorough and lengthy industry-wide effort would be required to create effective guidelines that would be both sanctioned by the ATSC and effectively practiced by the entire audio ecosystem, from content creation to emission. They immediately began their work.

An Initial Step: Using ITU-R BS.1770

In order to create effective loudness guidelines, a uniform measurement method was deemed essential, and considered to be the cornerstone of any loudness normalization solution. Given that ITU-R BS.1770 had been approved and published in July 2006 (following years of development), the group soon agreed to base its solution on such a worldwide, common reference for loudness measurement. This recommendation introduced a value identified in units of *LKFS* (**L**evel integrated over the time of the segment, using **K** frequency weighting—a new equalization curve developed for loudness measurement, reflecting human perception—and relative to **F**ull **S**cale digital level). A detailed description of ITU-R BS.1770 is presented in Chapter 5.18.

An experiment at USC was conducted in the fall of 2007 by using typical broadcast content and having a group of six award-winning mix engineers "normalize" the clips by ear in a calibrated listening environment. Fader information was collected from these adjustments and compared to the BS.1770 measured values for the same content. The maximum deviation between objective measurement data (via a BS.1770

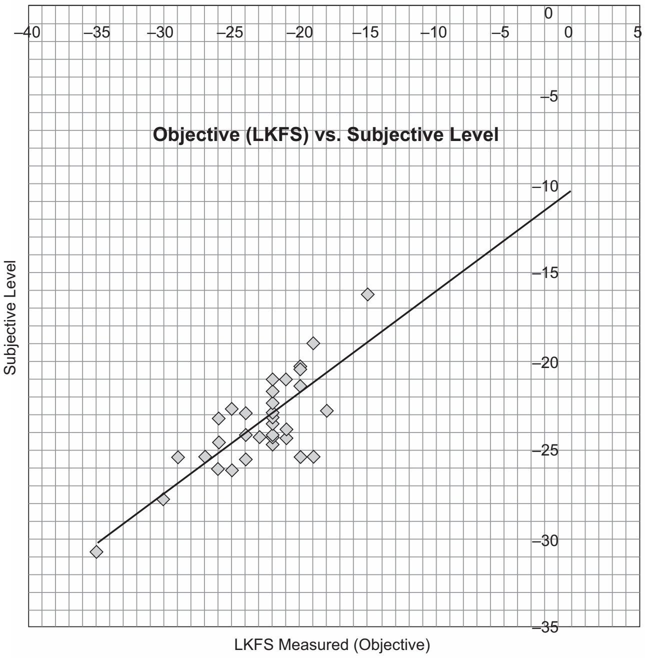

FIGURE 5.14-6 ATSC S6–3 objective versus subjective loudness measurement experiment results.

loudness meter) and subjective data (expert listener reaction) was less than +2 dB.[14] (See Figure 5.14-6.) With these firsthand results, the group was satisfied with the performance of the measurement recommendation and could now move on to consider additional essential elements of study for their main objective—a recommended practice on loudness.

ATSC A/85: Techniques for Establishing and Maintaining Audio Loudness for Digital Television

With group acceptance of the new ITU loudness measurement recommendation secured, S6–3 focused its work on the practices engineers would need to master and implement in order to solve the loudness problem. The initial draft for the group's recommended practice, to be called "ATSC A/85: Techniques for Establishing and Maintaining Audio Loudness for Digital Television," contained sections on loudness measurement, audio monitoring environment, metadata management considerations and dynamic range management. As their work progressed over the period from April 2007 to the initial release of A/85 in November 2009, followed by updates in 2011 and 2013, this list would expand as the group continuously tracked the loudness problem by reacting to ongoing technical activity in the industry and impending government regulation.

A/85 and the CALM Act

In 2008, the Commercial Advertisement Loudness Mitigation (CALM) Act was introduced in Congress as a reaction to consumer annoyance with the loudness of commercial advertising on television. In June of 2009, Congressional hearings were held and Congressional staff attended a technical demonstration on the progress of the ATSC's work, where it was shown that an industry-wide answer was near completion and its release was imminent. In October, a markup of the bill was created that would later become law in December 2010. The signed bill stated:

> Rulemaking Required. Within 1 year after the date of enactment of this Act, the Federal Communications Commission shall prescribe pursuant to the Communications Act of 1934 (47 U.S.C. 151 et seq.) a regulation that is limited to incorporating by reference and making mandatory (subject to any waivers the Commission may grant) the 'Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television' (A/85), and any successor thereto, approved by the Advanced Television Systems Committee, only insofar as such recommended practice concerns the transmission of commercial advertisements by a television broadcast station, cable operator, or other multichannel video programming distributor.[15]

The revised bill reflected a key shift away from subjective terms such as "strident" and "modulation" in favor of contemporary solutions, and this language ultimately became law. ATSC A/85's modern recommendations had now been validated as the solution to the irking problem of loud commercials on TV. These ATSC industry guidelines became government-mandated practices that TV broadcast stations and cable, satellite and telco TV operators must follow.

Six years later, after its effectiveness had been proven, ATSC A/85 was honored with a 2015 Academy of Television Arts and Sciences Engineering Emmy Award for dramatically improving consumer satisfaction.[16] See Figure 5.14-7.

FIGURE 5.14-7 Members of the ATSC Leadership and S6–3, receiving A/85's 2015 Academy of Television Arts and Sciences Engineering Emmy Award for dramatically improving consumer satisfaction, at the Bellagio Hotel, Las Vegas, NV, January 8, 2015. Left to right: Jerry Whitaker, VP, ATSC; Glenn Reitmeier, ATSC Board Chair; J. Patrick Waddell, ATSC S6 Chair; Jim Starzynski, ATSC S6–3 Chair; Mark Richer, ATSC President; Rich Chernock, ATSC TG-3 Chair; Greg Coppa, ATSC S6–3; Craig Todd, ATSC S6–3.

The next sections of this chapter will explain the mastering of A/85's vital technical practices for digital television audio loudness management when using the ATSC AC-3 audio system, including references to current government DTV loudness regulations and enforcement at this writing.

EFFECTIVE LOUDNESS NORMALIZATION USING ATSC A/85

Just like other important technical facets of television, the mastering of DTV loudness requires the harmonization of operational and engineering practices by the operator. This applies to the entire audio ecosystem from content creation, distribution and emission. There are a number of essential elements that must be considered that, when effectively coordinated and practiced, will yield an easily manageable loudness workflow for the professional, resulting in an enjoyable listening experience for the consumer.

These practices are documented and explained in ATSC A/85: Techniques for Establishing and Maintaining Audio Loudness for Digital Television. They are the basis of effective and compliant loudness control, mandated by the federal government and reflected in the FCC rules.

Applying the "Golden Rule"

The Golden Rule of the ATSC recommended practice is that the transmitted dialnorm value must correctly identify the loudness of the content it accompanies in order to prevent excessive loudness variation during content transitions on a channel (e.g., TV program to commercial) or when changing channels.[17] Implementation and consistency with this fundamental is the starting point and key to ensuring success and a good DTV loudness consumer experience.

To effectively apply the "Golden Rule," an operator must consider the following:

1. The importance of dialnorm.
2. The use of agile or fixed metadata during emission.
3. If the operator chooses a fixed metadata system that matches the chosen dialnorm value selected in the emission encoder, a content target loudness value must be universally implemented across the entire operation.
4. A content delivery specification should be written and distributed to all internal and external content suppliers explaining technical audio and loudness requirements and the expected characteristics of delivered content based on these needs. This specification should also describe any audio loudness practices used downstream by the recipient operator during content ingest, distribution and/or emission for reference by the supplier.
5. Good loudness metering practices ensuring the creation of compliant content must be employed at ingest, during production and postproduction, through distribution and emission, and for monitoring of short-form and long-form content. ATSC A/85 defines specific practices for these different operations.
6. Content creation, editing and monitoring facilities should be designed and equipped to adequately support operator loudness judgments in all environments.
7. The operator should analyze all content delivery paths and ensure a suitable loudness control method is employed for each, preventing the emission of non-normalized, noncompliant content to the audience, with a goal of maintaining audio quality.

The next part of this section will explain how these important points contribute to effective loudness management. ATSC A/85 documents the necessary processes. An explanation of each element and how they work together follows.

Dialnorm

During the initial design of the AC-3 system in the late 1980s and early 1990s, those responsible for its specification and features were very familiar with both current analog practices for loudness control and television audio soundtrack creation. As explained earlier, an analog broadcast processor protected against distortion and overmodulation, smoothed overall loudness for the listener and was effective at doing so over a broad range of content loudness at its input. As a consequence, however, the dynamic range of the content as delivered by the supplier was altered and no longer represented the full intent of the content creator. At the time, content came from many suppliers working in different genres at numerous facilities. Although reference setup tones on TV content were delivered at −20 dBFS for the most part, mixes were not delivered under specific dialog loudness standards. Those developing AC-3 recognized this and did not believe that this practice could change.

With this in mind, the AC-3 system was designed to manage loudness and dynamic range in a completely different manner than its analog predecessor. With a goal of carrying the content creator's intent all the way through to the listening audience, a means would need to be developed that could make use of digital's wide-range capabilities but still yield a pleasant listening experience for the audience. Dialnorm was invented to accomplish these goals.

The dialnorm concept:

- A mix engineer should have the latitude to choose how loud the dialog ("anchor") element is in a mix, placing it appropriately in the overall available dynamic range of the soundtrack, with genre-based considerations for soft passages, and allowance of headroom for loud effects and music.
- The *anchor element* is the perceptual loudness reference point or element around which other elements are balanced in producing the final mix

of the content, or that a reasonable viewer would focus on when setting the volume control.[18]

- Dialog forms the anchor element in the majority of television programs.[19] This is exemplified by a listener's choice to set his or her listening level to the storyline of the content. (A/85 describes proxies to be used as anchor elements if no dialog is present in the content.)
- With one common point identified as a "reference" or "anchor" section in any soundtrack, this point is loudness-measured and labeled with dialnorm metadata.
- Then, at content boundaries, *all elements* of every soundtrack are raised or lowered in level *by the system* once, so the anchors match each other in loudness during decoding at the listener's receiver.

Using this audio loudness management process makes the loudness of each soundtrack anchor perceptually the same, thereby normalizing loudness *and* meeting the goal of maintaining full creative dynamic range. This is essentially identical to a listener physically moving a volume control up or down at program and commercial boundaries to maintain a comfortable loudness, except this action is now happening automatically by using a continuous flow of changing dialnorm metadata in the bitstream to control the loudness of the receiver. Because *all elements* of the mix get louder and softer by the same increment as the anchor is adjusted up or down, dynamic range is unaltered.

Metadata System Options: Agile, Preset or Fixed

As an essential first step in an audio loudness management plan, the digital television operator must determine a metadata operating mode for the audio encoder. A/85 lists three possibilities: *agile metadata*, *preset metadata* and *fixed metadata*. The dialnorm parameter is the basis of the loudness system and is impacted by the choice of metadata mode selected. Many factors contribute to an operator's decision to choose one over the other. An explanation of how the dialnorm parameter is impacted by this choice follows.

Agile Metadata System

An agile metadata system (see Figure 5.14-8) is required to achieve the full capabilities of dialnorm. Because metadata must be present with the content and maintained at all points in the digital TV ecosystem, the following practices and conditions apply when a broadcaster chooses to pursue end-to-end dialnorm loudness implementation by means of an agile metadata system:

- Content is specified to contain a dialnorm metadata value encoded in the deliverable. The loudness value is left to the discretion of the content creator.
- The content recipient (operator) specifies how the dialnorm metadata will be encoded in the deliverable based on the systems the recipient has in place for ingesting and distribution of the content. A/85 specifies these techniques in its section 7.5.1. Two such methods are standardized by SMPTE.[20] The Dolby E codec is another method of delivery that also uses metadata encoding.
- The entire broadcast plant infrastructure must be compatible with synchronous and uninterrupted passage of the metadata stream.
- Monitoring equipment in the path must react to the dialnorm parameter in order to present an accurate representation of the content loudness that will be presented to the listener.
- The broadcast audio encoder must be capable of receiving the active dialnorm metadata and changing the dialnorm parameter in real time, under metadata stream direction.

Meeting the conditions needed for an agile system presents a difficult challenge to the broadcaster. This is based on the many different parts of the ecosystem that must work in perfect concert to rule out potential risk to content. If the flow of metadata becomes corrupted or interrupted for any reason, the operator should be aware that a protection or reversion mode engages in the broadcast encoder that must be provisioned to use a set of parameters suited to guarantee an acceptable (though not ideal) presentation of any supplied content.

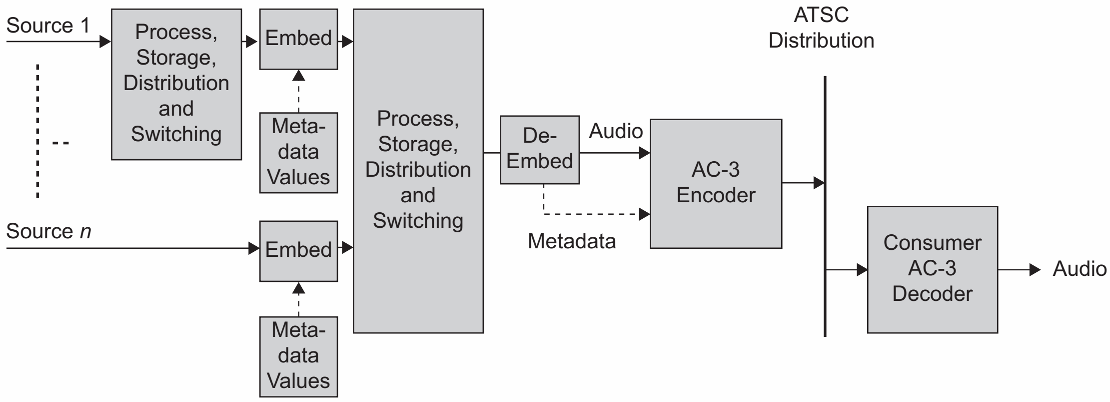

FIGURE 5.14-8 Simple diagram of an agile metadata system from ATSC A/85:2013.

As explained earlier, dialnorm is one of many metadata "control" parameters in the AC-3 system. Under fault conditions, an inaccurate dialnorm value of content loudness will make content soft or loud, but for the most part still understandable. However, if 5.1-channel surround sound content is mislabeled with incorrect channel metadata (also encoded as "control metadata" and carried in the same metadata stream as dialnorm), for example, 5.1 (3/2L mode) content is mistakenly labeled with stereo (2/0 mode) metadata, only the left and right channels of the 5.1 content will be heard. These channels usually contain only music and effects and in most cases do not contain dialog. In this situation without dialog, the listener is presented with an unacceptable condition, under any circumstance.

The conditions above, particularly along with the lack of interest by Hollywood in encoding metadata into a show master and with a resistance to the use of Dolby E by terrestrial broadcasters, make it difficult for a broadcaster to accept the risk associated with the complexity of an agile metadata system regardless of the benefit. Therefore, A/85 offers its other two alternatives, preset and fixed metadata.

Preset Metadata

The preset metadata system (see Figure 5.14-9) is based on engaging encoder presets loaded with different dialnorm values. This method is useable only when unique triggers are available from the plant automation system or by general-purpose interface (GPI) from master control.

If a category of content uses the same dialnorm value and plays out with a unique trigger, a preset can be engaged in the audio encoder with dialnorm metadata matching the loudness of this content. For example, a disc playback device stores long-form program inventory, all recorded at the same –27 LKFS dialog loudness, and when this device is selected by automation to go to air, the audio encoder receives a trigger from the automation system to engage a preset with a –27 dialnorm setting. The same station uses a different device to store short-form commercial inventory all recorded at the same –22 LKFS loudness. When this device is selected by automation to go to air, the audio encoder receives a trigger from the automation system to engage a different preset with a –22 dialnorm setting.

This technique allows for a change in dialnorm value based on content loudness but does not require the complexity of a fully agile system. It affords the operator some latitude in content loudness while reducing major risk. This method can also be used to switch other metadata parameters, for example, channel mode, potentially switching between 5.1 and two-channel stereo content. However, caution is required, since preset mode presents the possibility for an audible glitch in the receiver during switching, and therefore an AC-3 frame synchronizer should be considered to mitigate this problem.

Preset metadata mode was suitable early in the DTV transition, when content loudness and channel mode could be easily categorized. Programs were frequently delivered at the same dialog loudness and the same value worked well in many cases for the majority of two-channel commercials and promos. Considering the possibility for varied loudness across all content and its susceptibility to audible glitches, preset mode was limited in its usefulness, however.

Fixed Metadata

If one is able to imagine the notion of taking the agile metadata system and working in reverse, this would create the basis of the fixed metadata system (see Figure 5.14-10).

As described in the agile system, to fulfill the Golden Rule, dialnorm is set to match content loudness. In the fixed metadata system, to fulfill the Golden Rule, the opposite process is applied, by which *content loudness* is adjusted to match the dialnorm value. A single dialnorm value is set or "fixed" at the audio emission encoder, and all content is adjusted in level to meet the loudness indicated by this dialnorm setting.

To make this effective, content creators and broadcasters both must agree on a deliverable loudness recommendation that suits their needs. Broadcasters' main interest is minimizing risk to the content. A fixed system poses the least amount of such risk, having the audio encoder simply set to a value that is never changed. Content creators did not embrace the encoding of dialnorm metadata in their soundtracks, so they

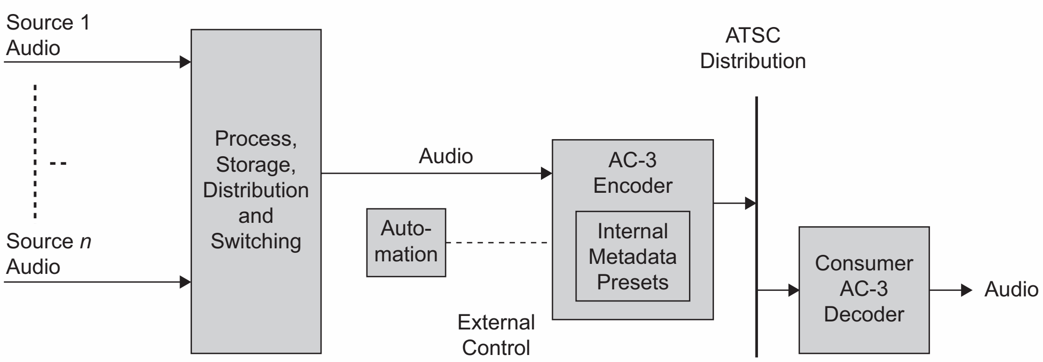

FIGURE 5.14-9 Simple diagram of a preset metadata system from ATSC A/85:2013.

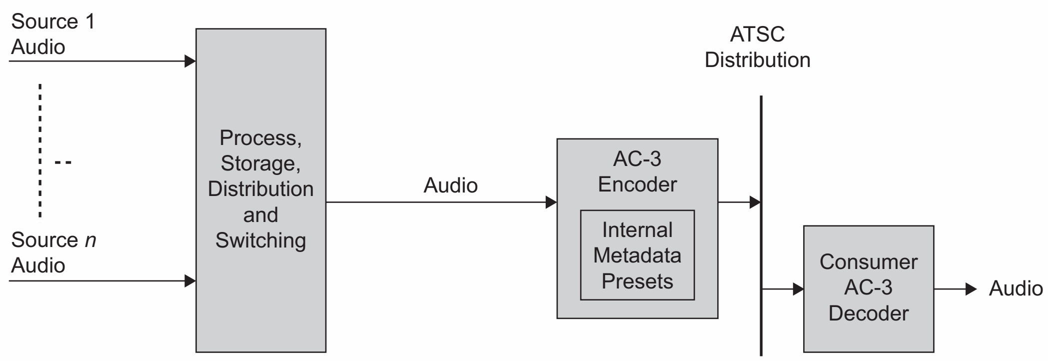

FIGURE 5.14-10 Simple diagram of a fixed metadata system from ATSC A/85:2013.

came to understand where the dialog loudness "sweet spot" for television could reside.

The Fixed Metadata System and ATSC Target Loudness Recommendation

The experts in ATSC S6–3, in conjunction with their content suppliers, agreed on the following industry recommendation for content dialog loudness, which was intended to work under a fixed metadata system:

> For delivery or exchange of content without metadata (and where there is no prior arrangement by the parties regarding loudness), the Target Loudness value should be −24 LKFS.[21]

As noted earlier, the creators of the AC-3 system never foresaw the possibility of this type of agreement between content creators and operators and therefore created the dialnorm system and the agile metadata concept in pursuit of a wide-range listener experience and creative flexibility during content creation. Nevertheless, a universal target loudness recommendation and the fixed metadata system became the preferred answer when operators looked for an alternative to agile dialnorm metadata.

The selected value of −24 LKFS was close to the figure for similar work going on in Europe by the European Broadcast Union (EBU), which had chosen −23 LUFS (Loudness Units Full Scale, an identical measure to LKFS). ATSC compromised at −24 LKFS, a figure that was also within reasonable reach of −27 LKFS, the average dialog loudness of typical episodic content at the time.

Given this agreed-upon recommendation, operators set their dialnorm values uniformly to −24 and specified content to match the dialnorm value, therefore fulfilling the Golden Rule and minimizing risk to content by reducing the complexity of the content delivery and broadcast distribution system. This undertaking resulted in a gradual improvement in program-to-interstitial transitions and established a uniform loudness across the dial. Broadcasters adopted the recommendation, and a significant improvement to TV loudness was presented to the listener.

Metadata System Choice

For all the reasons stated above, broadcasters should give serious consideration to developing individual workflows that support target loudness normalization of content upstream of the broadcast encoder, and the choice of a fixed metadata and dialnorm value. Use of a hierarchy of these workflows preserves quality and will be described later in this chapter. However, A/85 suggests the option of a loudness-control processor at the end of the signal chain to force target loudness at the output of the channel if upstream normalization of content is not practical. This choice should only be considered if no other means are available, as it limits dynamic range and alters the content, similar to applying range-altering NTSC techniques to DTV.

In most cases, the ATSC recommendation of −24 LKFS will be suitable to content creators and operators for exchange and emission of the long-form programs and events, short-form commercials, promotional material and public service announcements that make up most broadcast content.

Note that a fixed metadata system can work effectively using any value appropriate to the content as long as the Golden Rule of content loudness and dialnorm matching is fulfilled. Creators and distributors of "premium" content may choose a different average dialog loudness and dialnorm value, for example, −27 LKFS, to permit additional headroom for wider dynamic-range soundtracks.

CONTENT DELIVERY SPECIFICATIONS

A *content delivery specification* is required to identify the elements of good loudness practices.

Documenting and distributing the necessary technical requirements and recommendations for content delivery are important but often overlooked steps to managing audio at any facility. The purpose of the specification is to inform both *internal* and *external* content suppliers of the necessary steps to be taken when mixing audio, so the delivered soundtrack is compatible with systems used by the operator, and the soundtrack will present well to the audience. The specification strives to ensure that the content will meet the expectations of the creator, the operator and the audience.

ATSC A/85 is clear on a number of important guidelines that should be conveyed to those creating content and specified by the operator receiving the content. The operator should list these guidelines for the supplier, making certain to include any specifics that pertain to how the operator's handling of the content will impact the audio delivered to the listener.

Though other technical parameters for delivery should be included in the specification, this section describes the content delivery specification only as it applies to DTV loudness. Note that this information assumes the operator is employing a fixed metadata/dialnorm system. The specification should include the following:

- A reference to *ATSC A/85, Techniques for Establishing and Maintaining Audio Loudness for Digital Television*, for information on creating, measuring and distributing audio for DTV. It is a good practice to include a link to the ATSC website for access to the document, but not necessarily a direct link to the document, as the version may change and a direct link may become outdated.
- The use of ITU-R BS.1770–3 as the measurement method for determination of content loudness. (This Recommendation's *-3 revision* is required per ATSC A/85:2013 and FCC 14–71.)
- A Long-Form Content Loudness Specification: Specify the required *average* dialog loudness for soundtracks of long-form content. This value must match the dialnorm value that is carried in the operator's bitstream transmitted to the audience. It should be referred to in units of LKFS and be clear about any tolerance around the expected value, for example, "Program suppliers must provide a measured average of −24 LKFS (±2 dB) dialog loudness, not whispered or shouted." Note the extremes of the range are not to be targeted. Also include any special provisions that may be used by the broadcaster to correctly normalize any noncompliant content after it has been received. This may involve subjecting the content to file-based loudness scaling or real-time loudness processing, described later in this section, with its use at the discretion of the operator.
- A Short-Form Content Loudness Specification: Specify the required *average* loudness of the full program mix (not just the dialog element) for soundtracks of short-form content. This value must match the dialnorm value that is carried in the operator's bitstream transmitted to the audience. It should be referred to in units of LKFS, and be clear about any tolerance around the expected value, for example, "Program suppliers must provide a measured average of −24 LKFS (±2 dB) of the full program mix." Note the extremes of the range are not to be targeted. Include any special provisions that may be used by the broadcaster to correctly normalize any noncompliant content after it has been received. This may involve subjecting the content to file-based loudness scaling or real-time loudness processing, described later in this section, with its use at the discretion of the operator.
- Guidance on target value: Explain the correct interpretation of the numeric loudness value, and that the displayed value *will deviate* from the target average value due to the nature of audio dynamics.[22]
- An explanation on the use of wide dynamic-range, short-form content: State that the delivery of wide dynamic-range, short-form content may pose unintentional loudness attenuation of the anchor element within this type of content.[23]
- A True-Peak Specification: Specify a maximum true peak level for soundtracks.[24]
- A note on Metadata Authoring: State that the authoring of all DTV metadata will be by the operator/recipient. The operator should include a table specifying the in-use metadata parameters that will be transmitted with the content to audience receivers.
- An example of the loudness comfort zone: Consider including information pertaining to the loudness comfort zone (see Figure 5.14-11), and how use of the comfort zone impacts acceptable perceived loudness when transitioning between content items.[25]
- An explanation of downmixing: For operators accepting 5.1-channel soundtracks, note to the supplier that any 5.1-delivered soundtracks will be downmixed at time of broadcast for the stereo-listening audience in the viewer's receiver. The operator should note how the mixed elements of a 5.1 soundtrack will impact the stereo downmixed version with special consideration to loudness of the downmixed version.[26] (See Figure 5.14-12.)

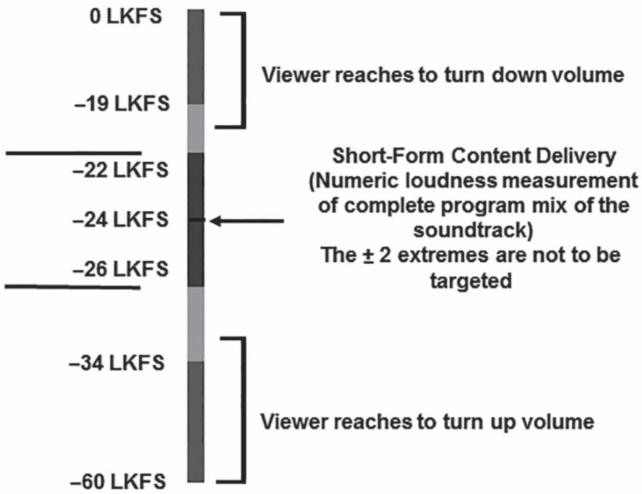

FIGURE 5.14-11 An interpretation of the loudness comfort zone using a −24 LKFS target for short-form content.[1] AuQ29

Understanding Elements in the Content Delivery Specification

Several key elements within the content delivery specification warrant fuller description for a thorough understanding by broadcasters:

1. *Target loudness and additional information on long- and short-form deliverables, their measurement, and how loudness is perceived:*

As described previously, when a fixed metadata system is chosen, a target loudness value must be established and enforced. This target must match the fixed dialnorm value in the metadata stream, thereby fulfilling the Golden Rule. There is a significant difference in measurement of long-form and short-form content described in A/85. Nevertheless, under a fixed system, a single loudness target is used effectively for both types of content for the following reasons:

- For long-form programming, a listener will set the volume control to the storyline of the content, which is usually normally spoken dialog—that is, not whispered or shouted—and this becomes the common "anchor element." Though the soundtrack may contain soft passages and loud effects, a listener will be tolerant of reasonable changes in range as long as they are of acceptable duration, and therefore listeners will not move the volume control under these conditions.
- Short-form content—for example, commercial advertising, promos, PSAs and other interstitial announcements—often focuses on overall perceptual impact to the listener, creating an overall anchor element (i.e., the entire mix serves as the anchor element), not a dialog/storyline-based anchor element. To accomplish this, these short-form elements are often produced with narrow dynamic range. Therefore, a full program mix measurement was chosen as the loudness determinant for these categories of content in A/85.

An example of the intended result presented to the listener:

> When a correctly mixed long-form program plays, the dialog level (not whispered or shouted) will be consistent from scene to scene and from act to act.[27] Loud effects and soft passages will be mixed with acceptable range and duration, referenced to the anchor dialog level. The listener sets the volume control to the storyline's dialog loudness, with the intention of setting it once and leaving it there. Next, a short-form advertisement with the identical average target value as the surrounding program (but measured using its full-program mix) plays during a commercial break in the program. With the volume control previously set for the *dialog* loudness of the program's storyline, the commercial's soundtrack transitions smoothly, based on the typical narrow dynamic-range characteristics of most ads and promos, and the use of the full program mix measurement.

2. *Wide-range short-form content poses a problem under the scenario described above*.

With a typical movie trailer as an example, if the trailer content is recorded with dialog (as many are), the wide-range music and effects will influence a significant portion of the average reported by the full program mix measurement. This will therefore reduce the perceptual loudness of any dialog in the trailer (due to loud music and effects), resulting in a poor transition from program to trailer. Because A/85 specifies a full program mix measurement for all short-form content, A/85 cautions the content producer on delivering wide dynamic-range content, and details the loudness results when doing so.[28]

An example of an unintentional result presented to the listener in such cases follows:

> When a correctly mixed long-form program plays, the dialog level will be consistent from scene to scene and act to act. Loud effects and soft passages will be mixed with acceptable range and duration, referenced to the anchor *dialog* level. The listener sets the volume control to the storyline's dialog loudness, with the intention of setting it once and leaving it there. Next, a short-form advertisement with a dialog element plays during the commercial break of the program. The short-form content has a wide overall dynamic-range containing effects and music that are loud relative to the dialog for a significant portion of the content. Although the short-form content has been properly delivered at the specified LKFS target *full program mix* value, the perceived loudness of the ad's *dialog* has been reduced well below the LKFS value of the adjacent program. The listener perceives the advertisement as softer than the program, due to the loudness mismatch of the dialog anchor loudness of program compared to the dialog loudness of the advertisement and its full program mix anchor.

Therefore, short-form content providers should be cautious when delivering DTV content with wide or extreme dynamic range. Mixes should always be influenced by knowledge of the features and the characteristics of the DTV system, as documented in A/85.

3. *The downmixing process and its impact on content loudness are sometimes misunderstood by content suppliers*.

When 5.1 content is specified and transmitted, the soundtrack presented to the portion of the audience that listens to the main soundtrack *in stereo* is created at the time of broadcast by the receiver, from the elements of the 5.1 soundtrack, under direction of operator-transmitted downmix control metadata accompanying the audio. (This excludes any separate alternate language and video description presentations.) See Figure 5.14-12.

If 5.1 content is mixed with identical phase and amplitudes of the same elements on multiple channels of the soundtrack, the resultant stereo downmix of the 5.1 content can be louder than the original 5.1 soundtrack.[29]

Therefore, content suppliers providing 5.1 channel soundtracks should measure both the loudness of the 5.1 version and the downmixed stereo version as created and monitored using the identical 5.1-to-stereo downmix metadata parameters used by the operator. The 5.1 mix should be adjusted to meet the required loudness for both versions.

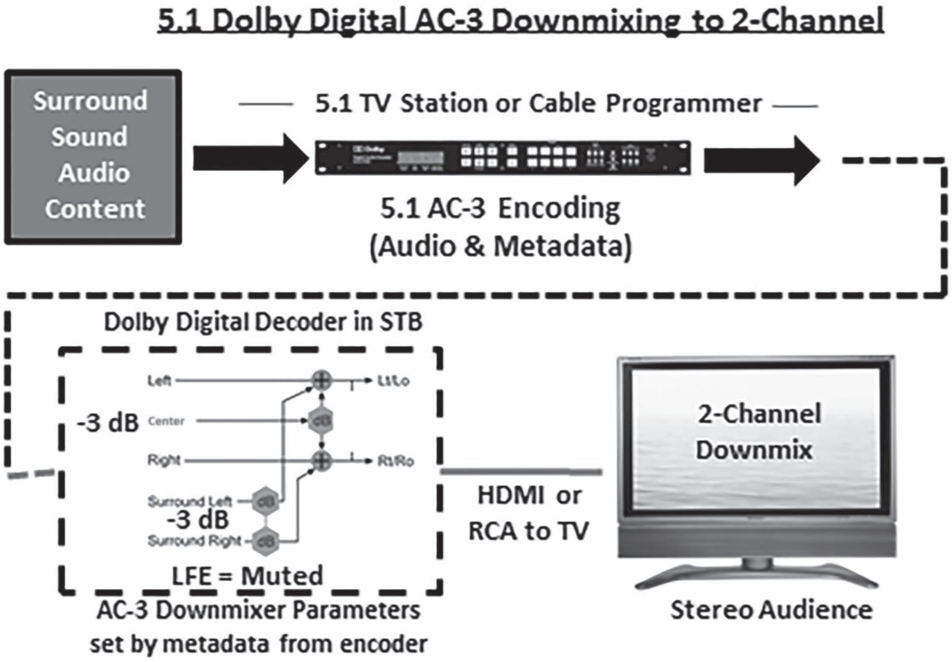

FIGURE 5.14-12 Typical ATSC 5.1-to-stereo downmix process.

A/85'S METHODS OF CONTROLLING PROGRAM-TO-INTERSTITIAL LOUDNESS

ATSC A/85's methods to control program-to-interstitial loudness are the key to all loudness normalization techniques. This section lists the choices for loudness normalization that apply to all types of content. Though this section may seem limited to loudness practices for content transitioning, the recommendations can apply to every situation for normalizing loudness at any step throughout the entire audio ecosystem. Therefore, these methods apply to workflows for live local and remote content creation, long- and short-form content ingest, postproduction and transmission operations.

With a fixed metadata system, accepted measurement practices and a target loudness recommendation in place that fulfills the Golden Rule, the operator can choose the management of loudness for all types and paths of content, one by one. These recommendations will be referred to in all parts of this section as they apply to various content creation and monitoring situations.

When audio quality preservation is a goal, these three methods of loudness normalization apply in this order of preference:

1. *Ensure that all content meets the target loudness, and that long-term loudness matches the dialnorm value.*
2. *Employ a file-based scaling device to match long-term loudness of non-conformant file-based content to the target value.*
3. *Employ a real-time loudness processing device to match the loudness of non-conformant real-time content to the target value.*[30]

The following two subsections detail each of the above methods.

Specifying and Verifying Loudness of Content as Supplied

When matching loudness to a target value is performed by the content creator, the highest-quality audio can be delivered to the audience by leaving loudness decisions in the hands of the supplier, thereby moving control of content loudness as far upstream as possible.

When working to a content delivery specification, normalizing audio as far up in the signal chain as possible, during the initial mix, permits content creators to establish the relationship of elements in the soundtrack based on their creative decisions, eliminating the need for further downstream normalization by the operator. Once the content supplier understands the necessary target required, content can be mixed and auditioned to meet expectations of the supplier and the recipient. (See Figure 5.14-13.) This practice will yield the highest-quality soundtrack to the audience if the distribution and transmission signal path is unaltered from creator to listener.

For an unaltered distribution path to be practical and for an operator to meet compliance, every item of content (e.g., locally and remotely originated live and recorded programs, ads, promos and postproduced material) must all be mixed and delivered meeting the loudness target.

If all content is mixed and delivered to a content loudness target value specification that is enforced

FIGURE 5.14-13 A contemporary audio console displaying a loudness histogram and an LKFS target value during content creation.

to match the dialnorm value, no other practices are required downstream to meet loudness compliance. Soundtracks presenting the supplier's unaltered creative intent can be delivered directly to the audience via a clean, process-free path. From a quality perspective, this is a best-case scenario.

The details of the necessary practices for this method of normalization will be described later in this chapter.

Taking Control of Content Loudness after Delivery

In many cases, especially for short-form content, some type of normalization may be required after delivery to the operator to meet compliance. A/85 specifies two means to do so, in this order of preference:

1. *Employ a file-based scaling device to match long-term loudness of non-conformant file-based content to the target value.*[31]

This method uses a transcoder type device that incorporates LKFS loudness normalization consisting of a "measure" and "scale" process. (See Figure 5.14-14.) The operator enters the target loudness value in the transcoder. On an automated first pass, applied file-based content is analyzed and a measured loudness value is determined. The device compares the measured loudness to the target loudness. Only if a difference is detected, during a second pass the scaler adjusts the overall loudness of all elements of the content to create a match.

Because the transcoder applies the loudness "scale" to all elements in a linear fashion, the dynamic range of the content is not altered, as long as there is enough headroom to scale louder (i.e., increasing gain) to the target without going into peak clipping. If there isn't, a limiter is required to prevent clipping. Loudness scaling that normalizes content softer (i.e., reducing gain)—which is typical for most short-form content—is simplistic, as it is reducing loudness with no need for limiting.

Because this operation is file based, it can be performed faster than real time (e.g., faster than playing linear tape), but the length of the process will vary based on content and manufacturer implementations.

Scaling is an excellent choice for the normalization of file-based content, as it is automated, precise and in most use cases will not alter the dynamic range or quality of the content. Most scaling devices work in the PCM domain and scale audio loudness of content to match the target. Other implementations may be available that can change the dialnorm value to match content loudness within AC-3 bitstreams, for use when agile and preset metadata systems might be chosen.

The focus of the loudness measurement during the "measure" and "scale" process must follow the rules of ATSC A/85. Remember that long-form content measurement is dialog based, while short-form content measurement is of the full program mix. The normalization section of the transcoding device must be specified and set up to perform the correct action on the content type. Some devices may include dialog recognition for long-form measure and scale, but some may not. In a rare instance, A/85:2013 footnotes the acceptable action of using ITU-R BS.1770–1 (formula 2) for automated devices.[32]

It should be noted that a manual scaling process is also possible. An operator can measure and scale content using an LKFS meter and adjust overall gain of the measured content to match a target. Because a unit of LKFS is the same size as a decibel, a loudness offset

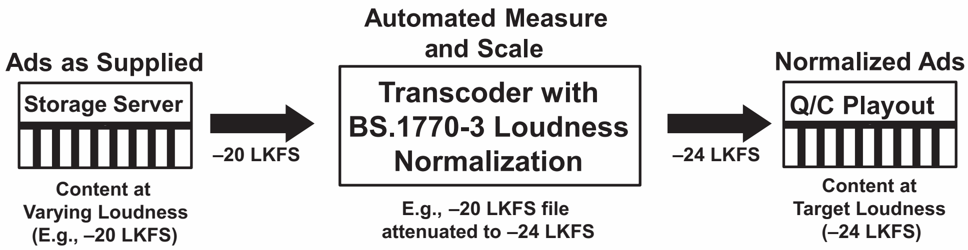

FIGURE 5.14-14 Simple diagram of automated measure and scale workflow.

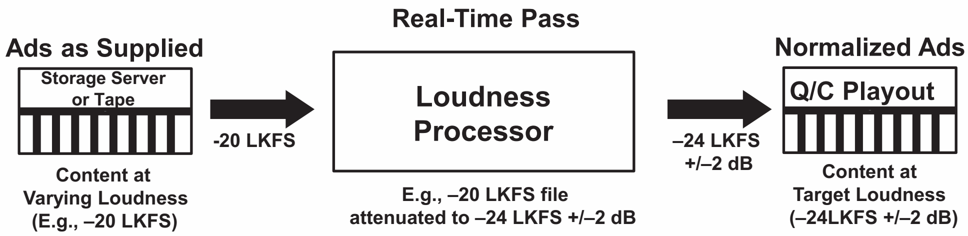

FIGURE 5.14-15 Simple diagram of loudness processing.

adjustment on the player device can be done in units of dB. Once the offset is determined and a unity gain copy of the content is made using the loudness offset applied at the player, the resulting tape or file will be correctly normalized.

2. *Employ a real-time loudness processing device to match the loudness of non-conformant real-time content to the target value.*[33]

This method uses a loudness-control processor to control range and fit content inside of a loudness target window. It is very effective for achieving content compliance by constraining the range of the applied audio in a similar fashion to the way an NTSC processor reduced range (and therefore altering the creative intent of the content) to prevent overmodulation, and as a by-product, making loudness transitions smoother by doing so.

If content either was not mixed to a target, or cannot be scaled (either with a file-based device or manually), the operator can employ a loudness processor to meet compliance. (See Figure 5.14-15.)

Note that the operator is well served by having a manufacturer's expert consult on the setup of the processor for every unique implementation. Though it may seem that a device can be set up quickly by choosing a preset named to fulfill a need, like "General Broadcast," a good match of processor to operation requires the setup be tailored to the type of content being normalized by the processor and the desired target. A processor with settings that are not well matched to content and loudness target can severely degrade the quality of the audio applied to it. This setup is best left to an expert on the specific device.

The simplest approach to overall audio loudness normalization is to apply a "one size fits all" processor downstream in transmission, at the output of the channel. However, if no normalization is enforced upstream, the unpredictability of the loudness levels of all types of content will require the processor to work very hard to place the varying loudness within a target window. Undesirable audible artifacts will most likely be created when doing so, which can be noticed by the audience and especially by content mixers and producers.

Hybrid Loudness Normalization Approach

With audio quality preservation as a goal, consideration should be given to establishing specific normalization workflows starting as far up in the signal path as possible. As it's unlikely that all content will be received or locally mixed to the loudness target, in many cases a plant-wide operation can benefit from use of a quality-driven, "hybrid" loudness normalization approach, using a combination of the methods described above. (See Figure 5.14-16.)

The benefit of using the hybrid approach is that the operator can make certain that quality is maintained whenever possible by considering the most preferred method of normalization, first based on content origin and available workflows, and moving down in preference to quality-altering methods only if necessary. (See Figure 5.14-17.)

For long-form content of various genres, enforcement of the content delivery specification should allow the broadcaster to use these programs with no need for further normalization, leaving loudness and creative decisions in the hands of the supplier. This applies to both externally supplied entertainment-type content as well as internally supplied news-type content, as examples. Working down the normalization

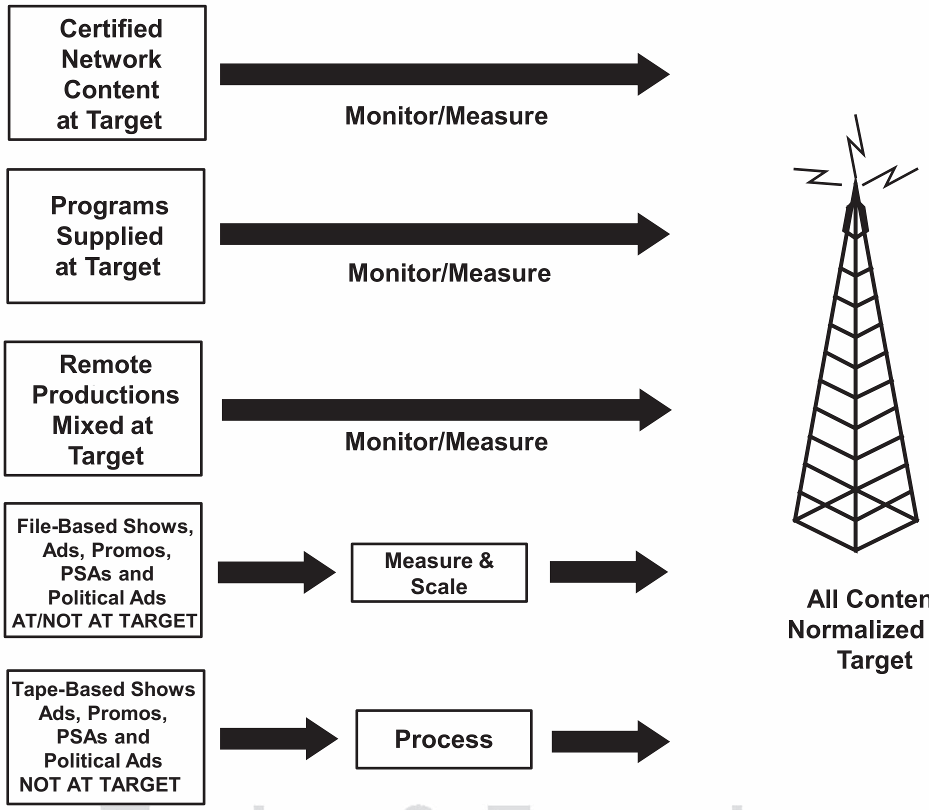

FIGURE 5.14-16 Simple diagram of the elements of a hybrid loudness normalization approach.

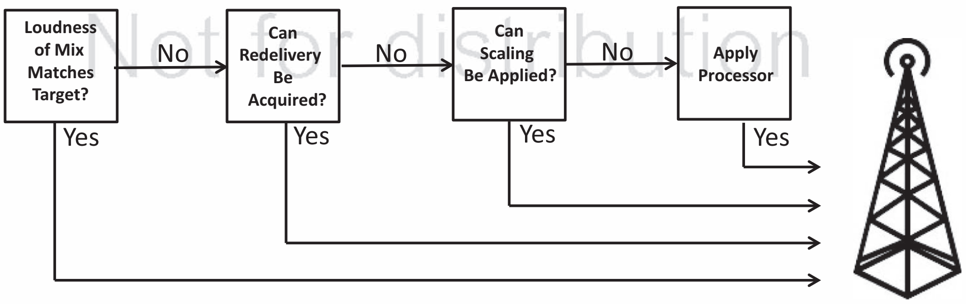

FIGURE 5.14-17 Order of preference for an audio-quality-driven loudness compliance process.

preference list, if this content cannot be relied on for loudness matching the target value and a replacement copy is not an option, the next best normalization choice is a file-based scaler with dialog detection for correcting a content loudness and target mismatch. If this is unavailable, next in preference, manual scaling can be performed. Last in preference, if manual scaling is not workable, the content can be normalized by a real-time processor.

For short-form content, enforcement of the content delivery specification *should* allow the broadcaster to air ads, promos, PSAs and other interstitials with no need for further normalization; however, it is unlikely that this process can always be relied upon. Advertisers and marketing departments strive to stand out among their competitors, frequently attempting to use a loudness advantage to do so. An operator can try to enforce the content delivery specification and ask for replacements when needed. However, the broadcast schedule, the sheer number of ads received and a delicate customer relationship may make this impractical. Based on this, a mandatory workflow that applies loudness scaling to all short-form, file-based content is a prudent step that should be considered by the operator.

If the operator sources file-based commercial advertising via a third-party aggregator, it is possible the aggregator may apply loudness normalization to content as a service to the recipient operator. The operator needs to determine if this is sufficient. If short-form content is acquired from other sources as well, it isn't practical to solely rely on the aggregator, and therefore it makes sense to normalize all short-form content through a scaling workflow managed by the operator.

If file-based scaling is not possible, or short-form content is received on tape, a manual scaling process can be used. If this is impractical, the ad or promo can be normalized by a processor.

With use of an effective hierarchical hybrid approach, the operator strives for a clean, unaltered path for compliant content to be presented to the audience without alteration. This is possible because the quality-altering processor is left only in content workflows that could not be normalized by more preferred methods upstream.

LOUDNESS MANAGEMENT DURING CONTENT CREATION

The first critical steps to identify when starting a project requiring the operator to meet a loudness target value are:

- Identify the content being mixed as long- or short-form type.
- Mix content focusing on average *dialog* loudness for long-form and on average *full program mix* loudness for short-form content.
- Read the content delivery specification and know the target loudness that's required. Content must be delivered to this loudness. The operator is obliged to have this specified loudness match the dialnorm value encoded in the AC-3 bitstream, therefore fulfilling the Golden Rule.

The following steps have been found to be effective for creating compliant mixes that should not require any further normalization by the recipient operator:

- Verify that the ITU-R BS.1770–3 loudness metering is set up correctly and in plain sight:
  - If available, set the meter up to "ATSC" if offered. The meter will now label loudness values in units of LKFS (ATSC) and not LUFS (EBU), though derived values for each are identical.
  - Set up the term. Many engineers use a very effective 10-second rolling average.
  - Next, make sure all channels of the soundtrack are being measured. If a 5.1 soundtrack is being mixed, some meters, for example, the Dolby LM100 mentioned earlier, will not be able to read all channels of audio without the content being encoded to 5.1 Dolby Digital or Dolby E, and the encoder connected to the input of the meter.
  - Do not engage dialog detection.
- Verify the monitoring system has been calibrated as described in a later section of this chapter. If a monitoring-level sound pressure level (SPL) that corresponds to the room size cannot be verified, at least make certain the main loudspeakers are each calibrated for equal amplitude at the mix engineer's listening position, and the low-frequency effects (LFE) channel has been set 4 dB louder than the other speakers.[34]

Live and Postproduced Long-Form Content

- Bring up the talent vocal on the console fader while the talent is speaking normally, not whispering or shouting. Raise or lower the fader, observing the metered loudness value. Park the fader where the loudness target is achieved.
- If mixing in an SPL-calibrated room, the monitor pot will be fixed at a predetermined volume and should be left there.
- If mixing in a non-SPL-calibrated room, the monitor pot should be adjusted while observing the target value on the loudness meter while talent is speaking normally and set to a comfortable volume for the mix engineer and not moved once set.
- With talent vocal at the target loudness and monitors set and fixed, the mix engineer blends the remaining elements of the soundtrack by ear, verifying the isolated dialog's loudness periodically and confirming the LKFS reading matches the target value and the mixer's perceived loudness. ATSC A/85 notes the use of an acceptable ±2 dB tolerance around this value, but it is clearly stated that *the extremes of this range are not to be targeted*.[35]
- The dialog loudness of correctly mixed content will be consistent within and across the acts of a program and from actor to actor.[36]
- Mostly by ear, the mix engineer determines the loudness relationship of the elements of the soundtrack, making certain they blend well against the loudness of the dialog anchor, noting that dialog intelligibility should never be inadvertently compromised by other elements of the mix.
- If creating 5.1 content, the two-channel downmix derived using the same audio encoder metadata parameters as the operator (typically −3 center, −3 surround and mute LFE) should be checked frequently during the mix to confirm mix and loudness compatibility of both versions.
- The mix engineer can refer to the loudness comfort zone for guidance in understanding when a listener will reach for the volume control to adjust for varying loudness.[37]
- If the program does not contain a dialog anchor, for example, in the case of a musical concert, or some types of documentary programming, a surrounding announcer vocal or a portion representing what a typical listener would set his or her volume control to, along with the loudness comfort zone, can be used as a reference to establish the loudness anchor of the non-dialog-based program.

These practices will yield good results that can be confirmed (if needed) if the live program is recorded and subsequently measured with a meter or software as follows:

- Using a manual approach, an operator can isolate the portions of the content where dialog (not whispered or shouted) is the only element present and take measurements in a spot-checking process by reviewing a few different sections of the content. The results should reveal that a correctly mixed program will have consistent dialog loudness that meets the target.
- In an automated approach, if the content is recorded as a file, use of a software program with dialog detection can scan the content faster than real time and report the average integrated dialog loudness value of the entire program.

Postproduced Short-Form Content

- Decide on the anchor element for the soundtrack, keeping in mind the measurement of the full program mix of short-form content will be used to determine the loudness value of all elements of the soundtrack, not just dialog.
- Bring up the anchor element (or multiple elements making up the anchor) on the console fader(s). Raise or lower the fader(s), observing the metered loudness value. Park the fader(s) where the loudness target is achieved.
- If mixing in an SPL-calibrated room, the monitor pot will be fixed at a predetermined volume and should be left there.
- If mixing in a non-SPL-calibrated room, the monitor pot should be adjusted while observing the target value on the loudness meter while talent is speaking normally and set to a comfortable volume for the mix engineer, and not moved once set.
- The mix engineer determines the dynamic range of the overall soundtrack and the loudness relationship of the elements within the soundtrack by ear. If dialog is included in the soundtrack, the listener will expect to be able to hear dialog normally. Like long-form content mixing, dialog intelligibility should never be inadvertently compromised by other elements of the mix.
- With the anchor element at the target loudness and monitors set and fixed, the mix engineer blends any remaining elements of the soundtrack by ear, verifying loudness periodically and confirming the LKFS reading matches the average target value and the mixer's perceived loudness.
- ATSC A/85 notes the use of an acceptable ±2 dB tolerance around the target.[38] However, it also notes that *the high and low side of this range should not be targeted*. If short-form content is delivered with an average value at the upper or lower side of this range, and the content recipient employs a scaling device to ensure loudness compliance, average loudness will be precisely readjusted to the recipient's target value specified in the scaling device. In these cases, delivering a value other than that specified by the content recipient may yield an unexpected, undesirable change in the overall perceptual loudness and composition of the soundtrack during air.
- To achieve desired results downstream, short-form content should be auditioned at the target loudness prior to delivery to the recipient. Mixing should be fine-tuned to make certain the full program mix meeting the target loudness value also meets the creative expectations of those producing the content.
- It is advisable that short-form content intended to play out directly after or before a program break and in a group with other short-form content (e.g., a commercial break composed of multiple commercials and promos) be auditioned playing side by side with dialog at the target loudness and with other typical short-form content at the target loudness, to hear the spot to be delivered in context.
- Noted above in this chapter, A/85 cautions about the use of wide dynamic range short-form content. Given its importance, that guidance is repeated here:

  > Those choosing to create and deliver wide dynamic range short-form content should note that the louder elements of this type of material will increase the loudness measured with a long-term integrated method, and consequently reduce the perceived Anchor Element loudness after normalization. This can cause an unacceptable match to long-form material measured with an anchor-based method.[39]

- If creating 5.1 content, the two-channel downmix derived using the same audio encoder metadata parameters as the operator (typically −3 dB center, −3 dB surround and mute LFE) should be checked frequently during the mix to confirm mix and loudness compatibility of both versions.

VERIFYING LOUDNESS

These practices will yield good results, but it is highly encouraged that the supplier verifies the loudness target of short-form content by subsequent measurement with a meter or software as follows:

- Using a manual approach, choose "infinite" or "integrated" measurement (continuous measurement, not short-term). Ensure the meter can measure all audio channels of the content. Disable any dialog detection. Clear the meter value, and pause the measurement. Play the content at the very beginning and precisely start the meter. Stop the meter at the exact end of the content. The resulting value is the average, integrated loudness value of the content used to report loudness per A/85.
- In an automated approach, if the content is recorded as a file, use of a software program measuring the full program mix can scan the content in faster than real time and report the average loudness value.

Ingest

For ingest practices, the operator can use similar methods to those described above to confirm that the loudness of content matches the target loudness, repeated here:

For Long-Form Content:

- Using a manual approach, an operator can isolate the portions of the content where dialog (not whispered or shouted) is the only element present and take measurements in a spot-checking process by reviewing a few sections of the content. The results should reveal that a correctly mixed program will have consistent dialog loudness and, if mixed as described above, that it meets the target.
- In an automated approach, if the content is recorded as a file, use of a software program with dialog detection can scan the content faster than real time and report the average integrated dialog loudness value.
- If a correctly set up automated scaler is employed, the scaler ensures that all average integrated content loudness meets the target value.

For Short-Form Content (these practices can be used to confirm commercial advertising loudness):

- Using a manual approach, choose "infinite" or "integrated" measurement (continuous, not short-term). Ensure the meter is measuring all audio channels of the content. Disable any dialog detection. Clear the meter value, and pause the measurement. Play the content at the very beginning and precisely start the meter. Stop the meter at the exact end of the content. The resulting value is the average, integrated loudness value of the content used to report loudness per A/85.
- In an automated approach, if the content is recorded as a file, use of a software program able to measure the full program mix can scan the content faster than real time and report the average loudness value.
- If a correctly set up automated scaler is employed, the scaler ensures that all content loudness meets the target value

Transmission or Master Control

The transmission engineer can verify the loudness of incoming programs from the studio or remote locations similar to the way other engineers verify the loudness of long-form content:

- Verify that the ITU-R BS.1770–3 loudness metering is set up correctly and in plain sight:
  - If available, set the meter up to "ATSC" if offered. The meter will now label loudness values in units of LKFS (ATSC) and not LUFS (EBU), though derived values for each are identical.
  - Set up the term. Many engineers use a very effective 10-second rolling average.
  - Next, make sure all channels of the soundtrack are being measured. If a 5.1 soundtrack is being mixed, some meters, for example, the Dolby LM100 mentioned earlier, will not be able to read all channels of audio without the content being encoded to 5.1 Dolby Digital or Dolby E, and the encoder connected to the input of the meter.
  - Do not engage dialog detection.
- Using conventional VU and not loudness metering, in a pre-program tech-fax check, have the studio or remote engineer send a 1 kHz tone to validate unity gain between locations and adjust as necessary. Acquisition of a steady −20 dBFS is typically used for this process.
- Next, have the studio or remote engineer send dialog (not whispered or shouted; spoken live by talent, a production assistant or from a recording) through the mix and distribution path that will be used for the program or event.
- Ensure the engineer confirms that the target loudness set on the mix console has been verified by loudness meter at the mix location that is set up in a similar fashion to the transmission engineer's loudness meter.
- The transmission engineer measures the content being sent by the studio or remote mix engineer, comparing values. If the distribution path is at unity gain and metering is set up correctly on both sides, the readings will be similar and the means to establish good loudness has been confirmed.
- Situations can vary and common sense should be used, but in cases where downstream monitoring has revealed a problem or discrepancy upstream, an accepted practice is to have the responsibility for correcting the problem as far upstream as possible. Once fixed upstream at the source, this reduces the potential for inadvertent re-correction of the problem downstream, only making matters worse.

The Monitoring Environment

An essential practice for making good loudness judgments and overall good audio mixing is working in a room with quality monitoring that has been calibrated for correct monitor gain, established in reference to room size. ATSC A/85 does a thorough job of explaining the requirements and the practices used to establish reliable environments capable of content creation meeting expectations and yielding interoperability across facilities.

Three approaches for room calibration are documented in A/85. The operator should choose the most comprehensive method practical for the operation in this order of preference:

1. The first approach is intended for planners, design engineers, installation engineers, maintenance engineers and users.[40] This is a comprehensive guide to the characteristics of rooms and spaces,

installation of the system and setup method to achieve calibration of monitoring. The key elements contained are two tables categorizing audio room types typically found in broadcast environments (see Table 5.14-2) and appropriate sound pressure levels for rooms based on room volume in cubic feet (see Table 5.14-3). Also included is a step-by-step method to calibrate a room in the correct fashion based on conditions from the two aforementioned tables.

2. The second approach is a quick reference on monitoring setup for television.[41] Its purpose is to describe just the necessary steps to ensure unity gain through the mix console along with appropriate test noise that can be played and monitors set for SPL, per the included room volume table.
3. The third approach is a really quick reference guide and is clever, simply instructing the user to set unity gain in the system and compare and adjust mixed content to a known item of content at −24 LKFS, the example target loudness, by ear.[42]

It should be noted that the embedded links in A/85 may not connect to available audio test content as documented in this section of the Recommended Practice. A/85 specifies the type of test content needed, and the operator should acquire suitable test material elsewhere if not available from ATSC.

In addition to room calibration, A/85 also gives guidance on room acoustics and loudspeaker placement, describing room layout and construction along with suitable acoustical treatments.[43] A section on room correction is included as well, explaining to the reader

TABLE 5.14-2
Categories of audio control rooms and spaces used in TV production (from A/85:2013).

Category	Characteristics
I	Principal audio monitoring control rooms with specialized acoustics and sound systems. Channel range up to 5.1 (3 front/2 surround/0.1 low-frequency enhancement). Well isolated from other operations. Widest frequency and dynamic ranges equal to best home cinemas properly aligned. This type of room may be used for quality control at the network level, for example, checking program material for conformance to delivery requirements when a question arises at ingest stations. Sound monitor quality dominates over production requirements in this category of room. Broadcast organizations might be expected to have only a small number of such rooms.
II	Audio-mostly production spaces with equipment needs and placement supplanting absolute audio monitoring conditions, although audio monitoring is still expected to be good. Channel number equal to highest number used for material originating in the room. Good isolation from other operations. This type of room may be used for program origination, with its output occasionally subject to check in a Category I room. Low-frequency range and headroom may be somewhat restricted compared to a Category I room.
III	Audio editing spaces, premix and prelay rooms, and other spaces the output of which is typically expected to be integrated into programs in a Category II room or better. If used for final mixing, apply the level and equalization recommended practice herein.
IV	Trucks and booths for program mixing. These spaces have special considerations due to their small room volume, high background noise level, high level of early reflections, and communication needs in a production environment.
V	Headphone monitoring systems recommendations. Used for ingest stations in crowded environments, quality control in machine rooms, and the like.

TABLE 5.14-3
Reference sound pressure level (SPL) recommendations for mixing rooms of varying size (from A/85:2013).

Categories	Room Volume in Cubic Feet	SPL in dB re 20 μN/m²
I, II	> 20,000	85*
	10,000 < 19,999	82
	5,000 < 9,999	80
	1,500 < 4,999	78
	< 1,499	76
III	Depends on room usage. For editing purposes, may be controlled by the editor for use with the material at hand. For final program mixing, follow the recommendations for categories I, II above.	
IV	< 1,500	76
V		Use 2 cc coupler and set 440 Hz level to 74 dB.

* Per SMPTE RP 200 [6]

both effective and ineffective methods to achieve good results.[44]

Practical room considerations:

- If possible, the audio mix room should be designed as an isolated acoustical space with high-quality loudspeakers that are suitable for the content being mixed and are comfortable and non-fatiguing for the mix engineer to work with.
- Most importantly, however, the monitor system, especially one for 5.1 surround sound, must be calibrated in some fashion, making it a reliable reference and not a handicap to accurate content creation. At minimum, using pink noise, the main amps/speakers should be set to equal amplitude at the listening position with the LFE set 4 dB louder, using a simple SPL meter (slow response, C-weighting) or at least by ear.
- A correctly set up broadcast loudness meter should be connected to the audio system, able to measure all channels of the content being mixed or monitored. The meter must be installed in clear sight of the mix engineer or operator.
- Large rooms with high reference SPL yield wide dynamic range mixes. Smaller rooms using a lower SPL reference yield more range-constrained mixes. Though the AC-3 system is technically capable of cinema-like sound with >100 dB of range, realistic, good-sounding DTV mixes have been done in smaller rooms with ~78 SPL. Typical episodic content for digital television is frequently mixed with these considerations in mind.
- If applicable, the content lay-back session (the process of rejoining audio to video after audio postproduction mixing) is an excellent opportunity to double-check loudness for matching the target value and the mix for compatibility and quality. A small room with lower SPL, equipped with a small surround system and/or stereo speakers, the speakers in a flat-panel TV, and a loudness meter or loudness software are perfect tools for examination of the 5.1 and derived two-channel mix before dub and delivery.
- In a transmission room situation where audio monitoring is limited to a two-channel, rack-mounted speaker-panel device, an operator can get a good sense of continuity and presence of elements in a mix. In these cases, however, surround sound content will be downmixed to two-channel for listening. Presence of signal can usually be observed on LED bar graphs to determine level, and some units may include LKFS loudness monitoring. Though not able to monitor full surround sound, the panel is useful for spot-checking and quality control (QC), and it is very effective for isolating and critiquing channels of audio on speakers and especially with headphones.

THE CALM ACT AND FCC RULES

This section identifies important considerations for establishing a loudness normalization plan as specified in the 2011 FCC Report and Order based on the 2009 and 2011 releases of ATSC A/85. The full FCC Report and Order should be referred to, as this section does not include all the information and footnotes from that document. A later part of this section will list minor changes to the rules from the FCC's Second Report and Order issued in response the 2013 release of ATSC A/85. Again, the full FCC Report and Order should be referred to, as this section also does not include all the information and footnotes from the Second R&O.

When researching the necessary steps to complete loudness normalization that complies with FCC rules, the broadcaster or operator is strongly encouraged to consult legal counsel for an interpretation of the law and for guidance with the execution of a suitable operating plan that fulfills all mandated requirements.

For an interpretation of how these rules apply to a TV station, an NAB report from December 19, 2011, is included for reference.

This section lists:

- What the CALM Act requires of the FCC
- FCC documents and their current location
- The starting date of CALM Act enforcement
- The FCC's response to the CALM Act
- FCC loudness management rules
- NAB TV TechCheck, December 19, 2011
- FCC response to the 2013 release of ATSC A/85
- FCC loudness management rules (2014)

CALM Act Requirements

Reacting to the DTV loudness problem and identifying ATSC A/85 as a solution as described here and in an earlier section of this chapter, the Commercial Advertisement Loudness Mitigation Act that became U.S. law in December 2010 and enforceable in 2012 required the following:

> Within 1 year after the date of enactment of this Act, the Federal Communications Commission shall prescribe pursuant to the Communications Act of 1934 (47 U.S.C. 151 et seq.) a regulation that is limited to incorporating by reference and making mandatory (subject to any waivers the Commission may grant) the 'Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television' (A/85), and any successor thereto, approved by the Advanced Television Systems Committee, only insofar as such recommended practice concerns the transmission of commercial advertisements by a television broadcast station, cable operator, or other multichannel video programming distributor.[45]

The FCC responded to the signing of the CALM Act into law by adopting a Notice of Proposed Rulemaking (www.gpo.gov/fdsys/pkg/FR-2011-06-03/html/2011-13822.htm). Comments on the proposal were addressed, and ultimately a first Report and Order (https://apps.fcc.gov/edocs_public/attachmatch/FCC-11-182A1.pdf) was issued on December 13, 2011.

A second Report and Order (https://apps.fcc.gov/edocs_public/attachmatch/FCC-14-71A1.pdf) based on the 2013 release of ATSC A/85 was released on June 4, 2014.

The First Report and Order became enforceable on December 13, 2012.[46]

The Second Report and Order became enforceable on June 4, 2015.[47]

In the First Report and Order, the FCC summarizes the CALM Act's requirements of the Commission as follows:

> The CALM Act directs the Commission to incorporate into its rules by reference and make mandatory a technical standard, developed by an industry standards development body, that is designed to prevent digital television commercial advertisements from being transmitted at louder volumes than the program material they accompany. As mandated by the statute, the rules apply to digital TV broadcasters, digital cable operators, and other digital multichannel video programming distributors ("MVPDs").
>
> The rules [we adopt today] are designed to protect viewers from excessively loud commercials and, at the same time, permit broadcasters and MVPDs to implement their obligations in a minimally burdensome manner. As described below, we will require broadcast stations and MVPDs to ensure that all commercials are transmitted to consumers at the appropriate loudness level in accordance with the industry standard.
>
> In the event of a pattern or trend of complaints, stations and MVPDs will be deemed in compliance with regard to their locally inserted commercials if they demonstrate that they use certain equipment in the ordinary course of business. For the embedded commercials that stationsvand MVPDs pass through from programmers, we also establish a "safe harbor" to demonstrate compliance through certifications and periodic testing [See Figure 5.14-18]. This regime will make compliance less burdensome for the industry while ensuring appropriate loudness for all commercials.[48]

The Report and Order also notes the following:

> The "Golden Rule" of the RP is that the dialnorm value must correctly identify the loudness of the content it accompanies in order to prevent excessive loudness variation during content transitions on a channel (e.g., TV program to commercial) or when changing channels. If the dialnorm value is correctly encoded—if it matches the loudness of the content, which depends in turn on accurate loudness measurements—the consumer's receiver will adjust the volume automatically to avoid spikes in loudness.
>
> In addition to requiring the Commission to incorporate the RP by reference, the CALM Act requires the Commission to incorporate by reference "any successor thereto."
>
> In addition, we adopt our tentative conclusion that "all stations/MVPDs and not only those using AC-3 audio systems" are subject to our rules.
>
> We conclude that the statute makes each station/MVPD responsible for compliance with the RP as incorporated by reference in our rules with regard to all commercials it transmits to consumers, including both those it inserts and those that are "embedded" in programming it receives from program suppliers.[49]

The NAB followed with this report on the CALM Act and the FCC Report and Order:

> On December 13, 2011, the Federal Communications Commission adopted an Order required by the Commercial Advertisement Loudness Mitigation (CALM) Act designed to prevent commercials from being louder than programs surrounding them.
>
> (see July 11, 2011, issue of *TV TechCheck*)
>
> The Order incorporates the entire ATSC Recommended Practice on *Techniques for Establishing and Maintaining Audio Loudness for Digital Television* (ATSC A/85) into the regulations and makes A/85 mandatory for managing relative loudness of commercials and programs. The FCC plans enforcement to be based upon patterns of public complaints, instead of an audit program.
>
> First and foremost, all television broadcast stations (and MVPDs) are deemed ultimately responsible for controlling the loudness of all commercials they transmit (with respect to the program segments before and after them). Stations are expected to directly (or indirectly) follow the recommendations in ATSC A/85 to measure and thereby control the loudness of the audio segments fed to the service's AC-3 encoder (regardless of its physical location). There are multiple approaches available to broadcasters that can mitigate their exposure to a Notice of Apparent Liability for loudness violations.
>
> Most stations will find that operating at a fixed loudness (as measured per ATSC A/85) and leaving the "dialnorm" setting constant in the AC-3 encoder will be the most practical implementation approach.
>
> Stations technically can choose a different fixed operating loudness value for each program source (virtual channel); however, the loudness of the network feed will establish that virtual channel's operating point. Further, since program providers have many distribution outlets, and typically do not want to deliver different levels to different outlets, expectations are that most sources will be delivered to stations at the

CALM ACT CERTIFICATION

This is to certify that:

1. As required by Section 76.607 of Title 47 of the Code of Federal Regulations, all commercial advertisements embedded in programs carried on **This Television Network** are in compliance with the loudness control practices contained in Advanced Television Systems Committee (ATSC) A/85: Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television ("ATSC A/85 Recommended Practice") at the point of distribution by **This Television Network** to authorized reception equipment of downstream Television Stations.
2. Compliance with the ATSC A/85 Recommended Practice is determined by **This Television Network** through the use of equipment and associated software that is installed, utilized and maintained in a commercially reasonable manner.

Executed this X day of X
By: Senior Vice President, Broadcast Operations

FIGURE 5.14-18 Example of a CALM Act Certificate.

recommended operating loudness of −24 LKFS (as measured with the ITU BS.1770–1 method). This value is then entered into the AC-3 encoder's "dialnorm" setting.

The rules for broadcasters will be in Part 73.682(e), and include six major sections. The first section covers compliance with ATSC A/85, as briefly described above. The remaining sections cover the following:

1. Commercials inserted by stations
2. Embedded commercials—safe harbor
3. Use of a real-time processor
4. Commercials locally inserted by a station's agent—safe harbor
5. Demonstrating actual compliance

For commercials inserted by stations (1, above), the rule reads:

A television broadcast station that installs, utilizes and maintains in a commercially reasonable manner the equipment and associated software to comply with ATSC A/85 shall be deemed in compliance with respect to locally inserted commercials, which for the purposes of this provision are commercial advertisements added to a programming stream by a station prior to or at the time of transmission to viewers. In order to be considered to have installed, utilized and maintained the equipment and associated software in a commercially reasonable manner, a television broadcast station must:

(i) install, maintain and utilize equipment to properly measure the loudness of the content and to ensure that the dialnorm metadata value correctly matches the loudness of the content when encoding the audio into AC-3 for transmitting the content to the consumer;
(ii) provide records showing the consistent and ongoing use of this equipment in the regular course of business and demonstrating that the equipment has undergone commercially reasonable periodic maintenance and testing to ensure its continued proper operation;
(iii) certify that it either has no actual knowledge of a violation of the ATSC A/85 RP, or that any violation of which it has become aware has been corrected promptly upon becoming aware of such a violation; and
(iv) certify that its own transmission equipment is not at fault for any pattern or trend of complaints.

For the "embedded commercials—safe harbor" situation (2, above), the rule is much more complex. The Order presents the FCC's determination that the CALM Act establishes stations as ultimately responsible for the loudness of all commercials broadcast—even those inserted upstream. In general, it calls for stations with annual receipts of more than $14 million to either obtain a certification from the upstream provider that it is A/85-compliant, or perform at least two annual tests on that provider's content to verify its loudness compliance. Either of these alternatives establishes a degree of protection ("safe harbor") against fines to the station in the event that excessively loud commercials are inserted upstream and cause complaints.

If the spot-check approach is taken, the first check must be completed by December 13, 2012. (Spot-checking is defined to include 24 hours of measurement and analysis of the audio loudness transmitted by the broadcast station. The Order includes various suggestions and requirements on how to implement the spot-checking process.) This section also includes a requirement to perform a 24-hour spot-check after a "pattern of complaints" result in an FCC inquiry, with a progressive escalation process that can lead to fines in the event of continued non-compliance.

The Order also provides the option (3, above) for a station's use of a properly maintained real-time audio processor, with record-keeping requirements for demonstrating its consistent and ongoing use.

Also contained in the Order (4, above) is the option to establish a safe harbor for the special case where commercials are locally inserted by a station's agent. Stations may demonstrate compliance by relying on the third-party local inserter's certification of compliance with ATSC A/85, conditional upon meeting the terms detailed in the Order's relevant subsection.

Finally, a station also may document actual compliance with ATSC A/85 with regard to any commercial advertisements that may become the subject of an inquiry, and certify that its own transmission equipment is not at fault for any such pattern or trend of complaints (5, above).

The Order further explains that if after a broadcaster informs an upstream source of a loudness issue, and it is not fixed in a timely fashion, that each station carrying that source's content will be subject to liability if the problem persists. The progressive test and report process outlined in the Order should provide incentive to the source to fix the problem, given that some stations might stop carrying the source's content to avoid financial exposure.

A streamlined financial hardship waiver for some of the above processes is available to small broadcast stations. A "small broadcast station" is defined for purposes of the streamlined waiver as either a station with no more than $14 million in annual receipts, or one that is located in television markets 150 to 210. Small broadcast stations must file for such waivers by 60 days prior to the effective date of the rules.

The following are excerpts from the FCC Second Report and Order on CALM, issued June 4, 2014:

2. The Commission's rules implementing the CALM Act, adopted December 13, 2011, require digital TV broadcasters, digital cable operators, satellite TV providers, and other digital MVPDs to ensure that the commercials they transmit to viewers comply with the television industry's 2011 ATSC A/85 Recommended Practice (RP), which describes how the industry can monitor and control the loudness level of digital TV programming. As mandated by the statute, the Commission incorporated into its rules by reference and made mandatory the 2011 ATSC A/85 RP. The rules took effect on December 13, 2012.

3. Section 2(a) of the CALM Act mandates that the Commission's rules incorporate by reference and make mandatory "any successor" to the RP. On March 12, 2013, the ATSC published a successor document to its 2011 A/85 RP. As described by the ATSC, the Successor RP applies an improved loudness measurement algorithm to conform to the International Telecommunication Union's (ITU) updated BS.1770 measurement algorithm, "BS.1770–3." BS.1770–3 employs "gating" that will exclude very quiet or silent passages of a commercial when calculating the average loudness of that commercial. Use of the new algorithm may reduce the volume of some commercials in certain circumstances. The Successor RP also contains other minor changes that do not affect our rules.

Additionally:

5. As required by the statute, we adopt the Successor RP and will incorporate it by reference into our rules. We also find, as we tentatively concluded in the *FNPRM*, that the only substantive change created by the Successor RP as it relates to our rules is the change to the measurement algorithm to conform to BS.1770–3. This finding is consistent with the ATSC's description of the Successor RP and is not disputed in the record. As a practical matter, this change seems to be designed to prevent advertisers from using silent passages to offset excessively loud passages when calculating the average loudness of program material. Thus, once this Successor RP is implemented, consumers may notice a modest decrease in the perceived loudness of certain commercials. This change is consistent with the type of updates that we believe Congress intended the Commission to incorporate in its rules by specifying in the CALM Act that the Commission shall make mandatory successor versions of the RP.

6. We adopt the proposal in the *FNPRM* to make the Successor RP mandatory as of June 4, 2015, one year from the release date of this Second Report and Order. NAB, the only commenter on this issue, supports this approach.

CONCLUSION

Confronted with unintentional consequences of a new high dynamic range audio technology, industry and government came together, recognizing DTV's early loudness problem and establishing government regulation that codified an industry-developed solution. With the timely development of the effective techniques explained in this chapter, a group of audio experts became uniquely empowered to influence quality-driven rules, ultimately solving the problem to the satisfaction of government, industry and, most importantly, the television audience.

This chapter describes the story, solutions and successes of the first generation of digital television loudness control. Next-generation audio systems are already beginning to benefit from the pioneering efforts described here, leveraging and perfecting these practices that laid the groundwork.

ACKNOWLEDGEMENTS

The author wishes to thank his colleagues Margaret Tobey, Tomlinson Holman and Glenn Reitmeier for their assistance in compiling this chapter.

Notes

[1] Tim Carroll and Jeffrey Riedmiller, *NAB Engineering Handbook*, 10th edition, "Audio for Digital Television" (p. 1329), 2007.
[2] ATSC A/85:2013, Section 1.
[3] FCC 11-182A1, Report and Order, December 13, 2011.
[4] ATSC A/85:2013, Section 4.
[5] Dolby LM100 Broadcast Loudness Meter User's Manual, Issue 5.
[6] Mezzanine-level compression is used for efficient distribution in applications prior to final emission coding.
[7] Codec = Coder/Decoder.
[8] SMPTE ST 2020 Parts 1, 2, 3.
[9] Dolby E's eight channels could be assigned with six channels for a 5.1-channel surround mix, and the remaining two channels used for a synchronized stereo mix.
[10] AES3–2009, "AES standard for digital audio: Digital input-output interfacing—Serial transmission format for two-channel linearly-represented digital audio data," Audio Engineering Society, New York, reaffirmed 2014. www.aes.org/publications/standards/search.cfm?docID=13
[11] ATSC A/53 Part D 2005.
[12] ATSC A/85:2013, Section 10.1.
[13] ATSC S6–3–001r0.
[14] ATSC S6–3 Loudness Experiment Report, September 2009.
[15] S. 2847 (111th), CALM Act.
[16] ATAS Press Release, October 14, 2014.
[17] Federal Register, "Rules and Regulations," Vol. No. 131 (p. 40279), July 9, 2012.
[18] ATSC A/85:2013, Section 3.4.
[19] ATSC A/85:2013, Section 5.2.
[20] SMPTE ST 2020, Parts 1, 2 and 3.
[21] ATSC A/85:2013, Section 6.
[22] ATSC A/85:2013, Section 5.2.3.
[23] ATSC A/85:2013, Section 5.2.4.
[24] ATSC A/85:2013, Section 6.
[25] ATSC A/85:2013, Annex E.
[26] ATSC A/85:2013, Section 5.2.5.
[27] DTV Audio Group Meeting, AES Convention, Los Angeles, October 1, 2016.
[28] ATSC A/85:2013, Section 5.2.4.
[29] ATSC A/85:2013, Section 5.2.4.
[30] ATSC A/85:2013, Section 8.1.1.
[31] ATSC A/85:2013, Section 8.1.1.
[32] ATSC A/85:2013, Page 16, Footnote 3.
[33] ATSC A/85:2013, Section 8.1.1.
[34] ATSC A/85:2013, Section 10.3.
[35] ATSC A/85:2013, Section 6.
[36] DTV Audio Group Meeting, AES Convention, Los Angeles, October 1, 2016.
[37] ATSC A/85:2013, Annex E.
[38] ATSC A/85:2013, Section 6.
[39] ATSC A/85:2013, Section 5.2.4.
[40] ATSC A/85:2013, Section 10.
[41] ATSC A/85:2013, Annex D Section 1.
[42] ATSC A/85:2013, Annex D Section 2.
[43] ATSC A/85:2013, Annex B.
[44] ATSC A/85:2013, Annex C.
[45] S. 2847 (111th), CALM Act.
[46] FCC 11–182 December 13, 2011.
[47] FCC 14–71 June 4, 2014.
[48] FCC 11–182 December 13, 2011.
[49] FCC 11–182 December 13, 2011.